AP STAT: CHAPTER 3 **CATEGORICAL DATA**

MAKE A PICTURE!

First, create a frequency table

Example: number of students at CB South in each grade:

Grade	TOTAL
10	534
11	552
12	515

Proportion = decimal

Percent = %

Frequency = # of things

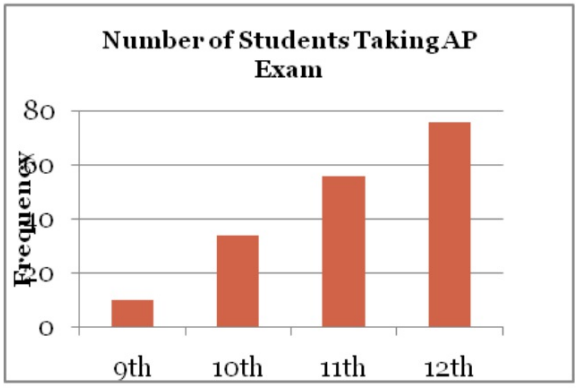
Relative frequency = % of things

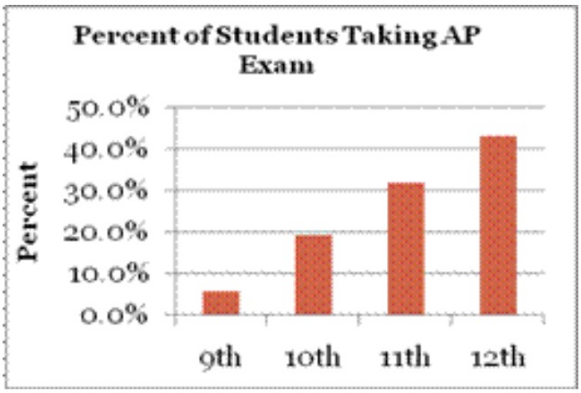
Distribution (of a variable)- shows values of the variable & how often the sample takes each value

Examples: Bar chart, pie chart, histogram, stemplot, etc.

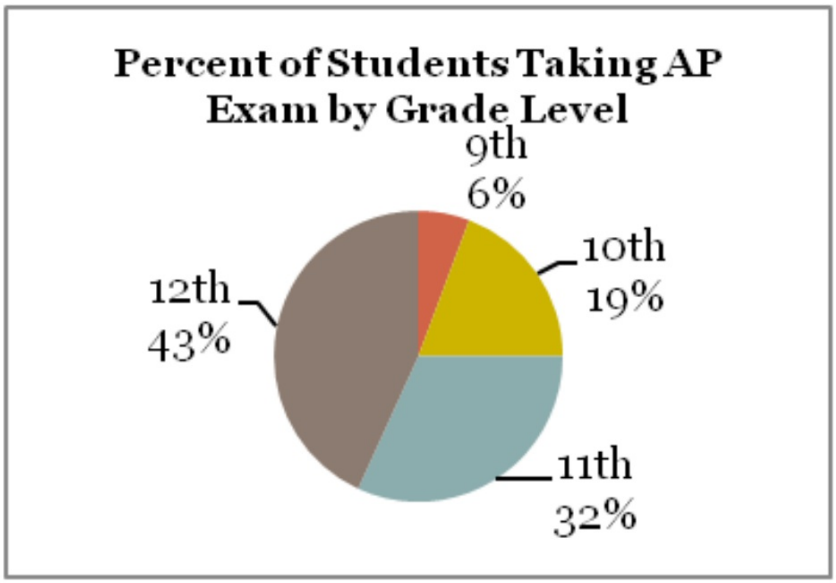
Categorical Distributions:

1. Bar Chart





Pie charts


Percent of Students Taking AP Exam by Grade Level
9th 6%
10th 19%
11th 32%
12th 43%

3. Contingency tables (aka Two-Way tables)

	Frosh	Soph	Junior	Senior	Total
Male					
Female					
Total					

Identify:

- Row variable = GENDER
- Column variable = GRADE
- Values of the variable = the different rows/columns
- Total (n) = bottom right of chart
- # of Cells (doesn't count totals)
- Totals (margins)

Example: Hospitals

	Hospital A	Hospital B
Died	63	16
Survived	2037	784

* What percent of people died?

* **Of** those people that went to Hospital A, what percent died?

* **Of** those people who went to Hospital B, what percent died?

* **Of** those people who died, what percent went to Hospital A?

* What percent of people died **and** went to Hospital B?

2 types of Distributions for Categorical Variables

1) MARGINAL DISTRIBUTIONS

· How to make: convert totals into %'s

· *Example: Hair color vs. Gender*

	Brown	Blonde	Black	Red	Total
MALE	26	24	10	3	63
FEMALE	20	35	12	6	73
TOTALs	46	59	22	9	136

* Marginal distrib. of HAIR COLOR:

* Marginal distribution of GENDER:

* Visually: BAR GRAPH

2) CONDITIONAL DISTRIBUTIONS

- Look at … one variable
- Then look at … each value of the variable
- Break down … each value into its pieces
- ALWAYS … in %
- Example: Hair Color vs. Gender

Find the conditional distrib. of HAIR COLOR:

	Brown	Blonde	Black	Red	Total
MALE	26	24	10	3	63
FEMALE	20	35	12	6	73
TOTALs	46	59	22	9	136

SEGMENTED BAR CHART:

Find the conditional distrib. of GENDER:

	Brown	Blonde	Black	Red	Total
MALE	26	24	10	3	63
FEMALE	20	35	12	6	73
TOTALs	46	59	22	9	136

SEGMENTED BAR CHART:

Try worksheet 3A on your own!
Feel free to work with someone else

AP Stat- worksheet 3A- Categorical Variables practice

In a survey of adult Americans, people were asked to indicate their **age** and to categorize their **political preference** (liberal, moderate, conservative). The results are as follows:

	Liberal	Moderate	Conservative	*Total*
under 30	83	140	73	*296*
30 - 50	119	280	161	*560*
over 50	88	284	214	*586*
total	*290*	*704*	*448*	*1442*

1. What are the row and column variables?
2. What percent of Liberals are under 30?
3. Of those over 50, what percent are Liberals?
4. Of those that are moderates, what percent are 30-50?
5. What percent of respondents are moderate and under 30?
6. Calculate the ***marginal distribution*** for the ***AGE variable***. Write these down. Then make a bar graph of the marginal distribution for age.

7. Calculate the ***marginal distribution*** for the ***PREFERENCE variable***. Write these down. Then make a bar graph of this marginal distribution.

8. Calculate the ***conditional distribution*** of the ***AGE variable***. Write these down. Then make a segmented bar graph of this marginal distribution.

9. Calculate the ***conditional distribution*** of the ***PREFERENCE variable***. Write these down. Then make a segmented bar graph of this marginal distribution.

1) Row var = AGE Column var = POLITICAL PREFERENCE

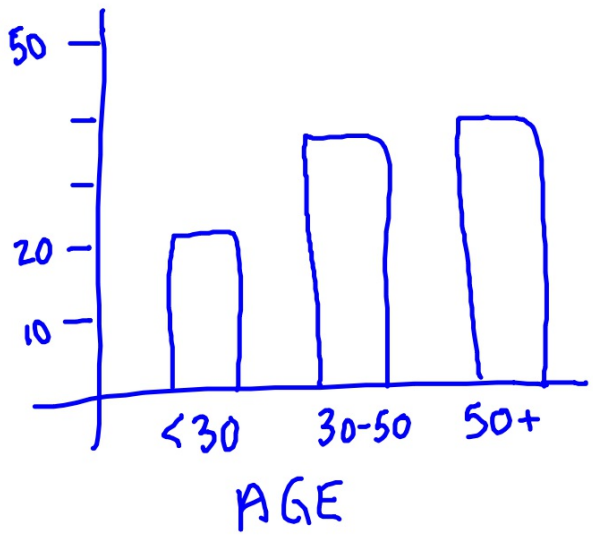
2) P(<30|L) = 83/290 = 28.6%

3) P(L|50+) = 88/586 = 15%

4) P(30-50|M) = 280/704 = 39.8%

5) P(M ∩ <30) = 140/1442 = 9.7%

6) <u>AGE</u>
<30 = 20.53%
30-50 = 38.83%
50+ = 40.64%

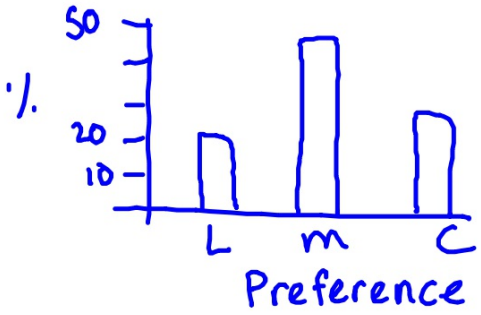


7) PREFERENCE

L = 20.11%

M = 48.82%

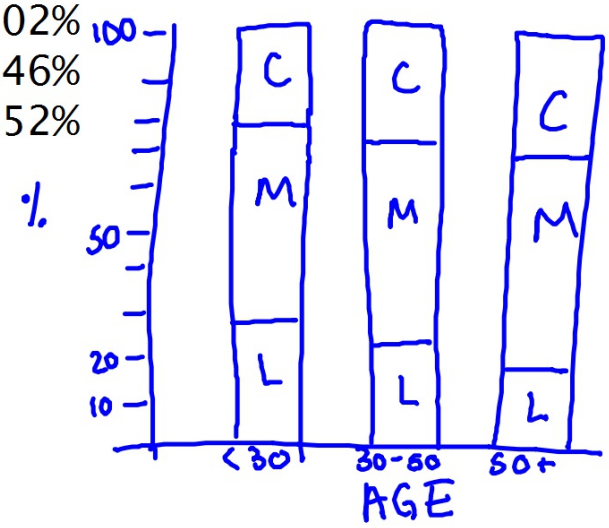
C = 31.07%



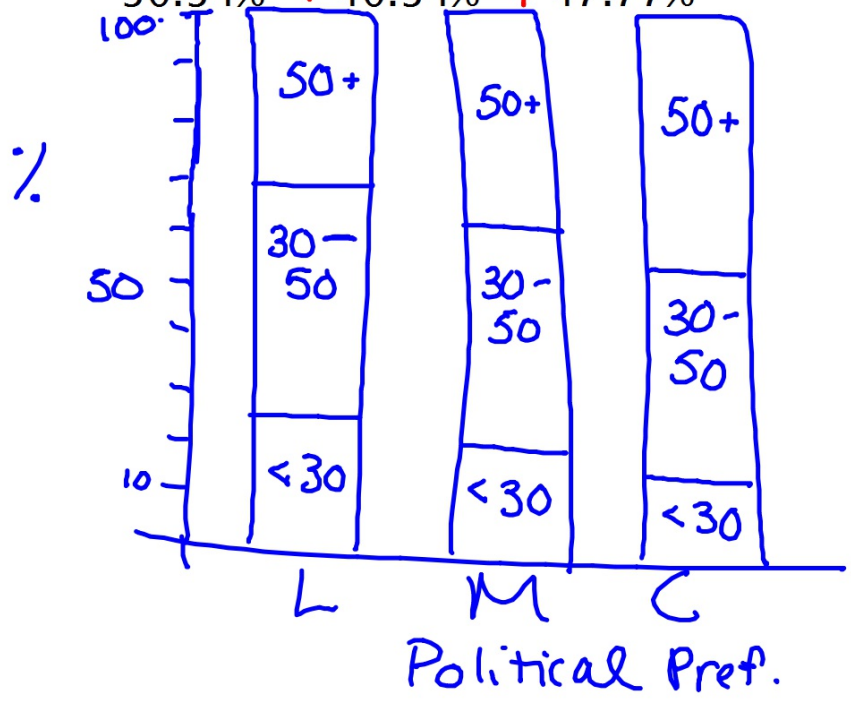
dependent

8)

	≤30	30–50	50+
L	28.04%	21.25%	15.02%
M	47.3%	50%	48.46%
C	24.66%	28.75%	36.52%



9)	L	M	C
<30	28.62%	19.89%	16.29%
30–50	41.03%	39.77%	35.94%
50+	30.34%	40.34%	47.77%



AP Stat- worksheet 3B- Categorical Variable practice

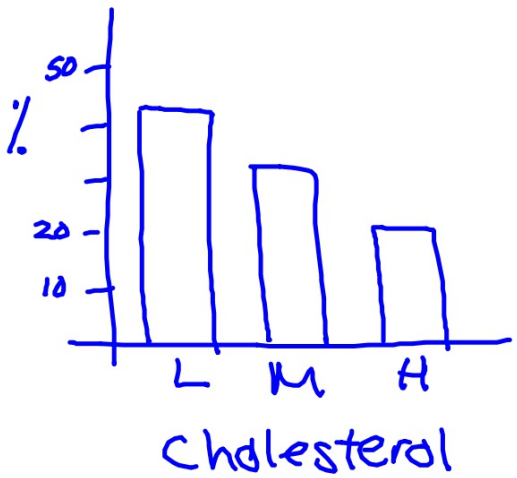
A 4-year study reported in *The New York Times*, on men more than 70 years old analyzed blood cholesterol and noted how many men with different cholesterol levels suffered nonfatal or fatal heart attacks.

	Low cholesterol	Medium cholesterol	High cholesterol	
Nonfatal heart attacks	29	17	18	64
Fatal heart attacks	19	20	9	48
	48	37	27	112

a. Calculate the marginal distribution for cholesterol level and make a bar graph.
b. Calculate the marginal distribution for severity of heart attack and make a bar graph.
c. Calculate three conditional distributions for the three levels of cholesterol and make a stacked bar graph.
d. Calculate the conditional distributions for the type of heart attack and make a stacked bar graph.

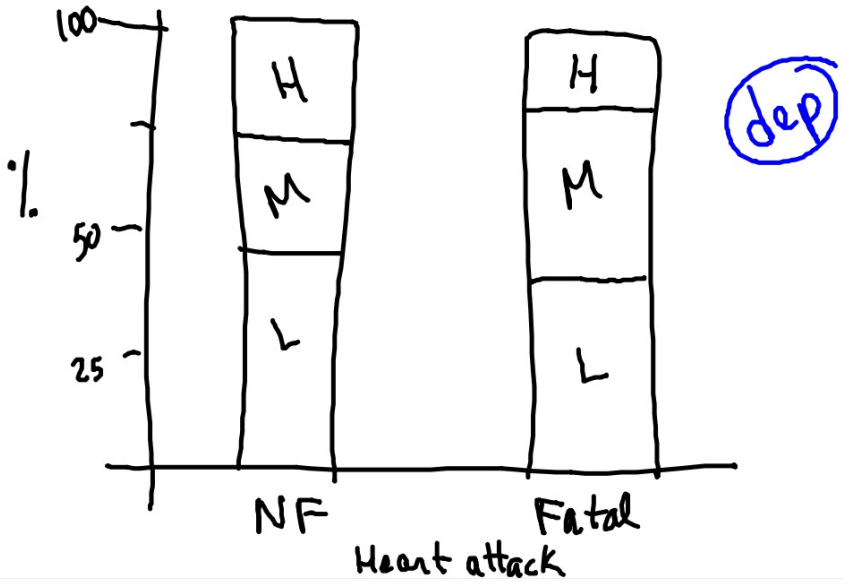
Answers to worksheet 3B:

(a) low: 42.9%
med: 33.0%
high: 24.1%



(d)

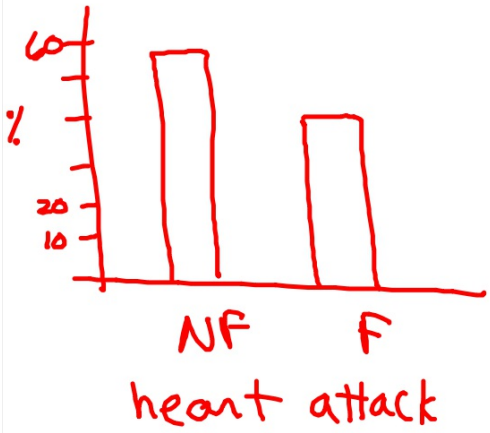
	NF	Fatal
Low:	45.3%	39.6%
Med:	26.6%	41.7%
High:	28.1%	18.8%



Worksheet 3B continued:

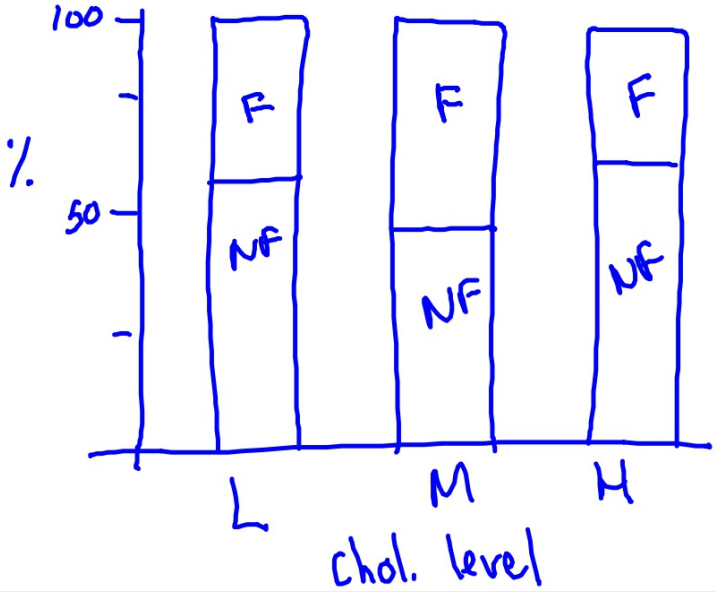
(b) NF: 57.1%

F: 42.9%



(c)

	low	med	high
NF	60.4%	45.9%	66.7%
F	39.6%	54.1%	33.3%



Worksheet 3C:

A study was made to compare year in high school with preference for vanilla or chocolate ice cream with the following results.

	Vanilla	Chocolate	
Freshman	20	10	30
Sophomore	24	12	36
Junior	18	9	27
Senior	22	11	33
	84	42	126

a. Calculate the marginal distribution for year in school

b. Calculate the marginal distribution for flavor preference

c. Calculate the conditional distributions for each year in school, and make a segmented bar graph. ~~What do you notice about the picture~~?

d. Calculate the conditional distributions for of the two flavors and make a segmented bar graph. ~~What do you notice about the picture?~~

Worksheet 3C:

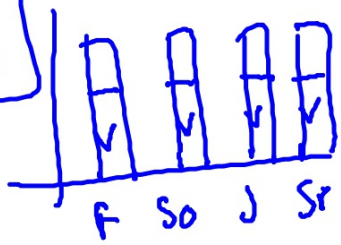
A study was made to compare year in high school with preference for vanilla or chocolate ice cream with the following results.

	Vanilla	Chocolate
Freshman	20	10
Sophomore	24	12
Junior	18	9
Senior	22	11

Indep

(a) Marginal for grade level:
Fr: 23.8%
So: 28.57%
Jr: 21.43%
Sr: 26.19%

(b) Marginal for flavor:
Vanilla: 66.7%
Chocolate: 33.33%



(c) Conditonal for grade level:

FROSH	SOPH	JR	SENIOR
V: 66.67%	V: 66.67%	V: 66.67%	V: 66.67%
C: 33.33%	C: 33.33%	C: 33.33%	C: 33.33%

(d) Conditonal fo flavor

Vanilla	Chocolate
Fr: 23.81%	Fr: 23.81%
So: 28.57%	So: 28.57%
Jr: 21.43%	Jr: 21.43%
Sr: 26.19%	Sr: 26.19%

Independence: when one variable does not affect another variable.

aka: NO association btw. the 2 variables

NO relationship

How to tell?

- When the marginal distr. of one variable = conditional distr. of other variable
- When stacked bar graph looks same throughout all values of variable *

Worksheet 3c: Question #2

2) An organization is concerned about the number of new employees that leave the company before they finish a year of work. So they predict this! Below are the results:

	Actually stay	Actually leave
Predicted to stay	63	12
Predicted to leave	21	4

Are the two variables independent? Justify your answer!

Worksheet 3C: question 2

Marginal for Prediction:
Pred. stay: 75%
Pred. leave: 25%

Conditonal for Actual:

actual stay:	***actual leave:***
Pred. stay: 75%	pred. stay: 75%
Pred. leave: 25%	pred. leave: 25%

Marginal for actual:
Actual stay: 84%
Actual leave: 16%

Conditonal for predicted:

Pred. stay	***Pred. leave***
act. stay: 84%	act. stay: 84%
act. leave: 16%	act. leave: 16%

Start on Classwork!